Accelerated Image Reconstruction using Ordered Subsets of Projection Data

H. Malcolm Hudson, Richard S. Larkin

Abstract— **We define ordered subset processing for standard algorithms (such as Expectation Maximization, EM) for image restoration from projections. Ordered subsets methods group projection data into an ordered sequence of subsets (or blocks). An iteration of ordered subsets EM is defined as a single pass through all the subsets, in each subset using the current estimate to initialise application of EM with that data subset.**

This approach is similar in concept to block-Kaczmarz methods introduced by Eggermont et al [1] for iterative reconstruction. Simultaneous iterative reconstruction (SIRT) and multiplicative algebraic reconstruction (MART) techniques are well known special cases. Ordered subsets EM (OS-EM) provides a restoration imposing a natural positivity condition and with close links to the EM algorithm.

OS-EM is applicable in both single photon (SPECT) and positron emission tomography (PET). In simulation studies in SPECT the OS-EM algorithm provides an order-of-magnitude acceleration over EM, with restoration quality maintained.

Keywords— **EM algorithm, emission tomography, MART, OSL algorithm.**

I. Background

The application of Expectation Maximization (EM) algorithms in emission tomography by Shepp and Vardi [2] has led to the introduction of many related techniques. These include a number of important Bayesian (or equivalently penalized likelihood) adaptions of EM. Recent papers have emphasized the quality of the reconstruction offered by these algorithms. See Hebert and Leahy [3] and Green [4] for some recent Bayesian developments, and Chornboy et al [5] for an evaluation of the benefits of EM in single photon emission tomography (SPECT).

While quality of reconstruction is good, the application of EM is computer intensive, and convergence slow, even with standard acceleration techniques (Kaufman, [6]). We provide here an ordered subset (OS) algorithm that processes the data in subsets (blocks) within each iteration and show that this procedure accelerates convergence by a factor proportional to the number of subsets. In SPECT, the sequential processing of ordered subsets is very natural, as projection data is collected separately for each projection angle (as a camera rotates around the patient in SPECT); counts on single projections can form successive subsets.

Computational benefit from such 'divide and conquer' processing may be anticipated, as in applications in sorting and fast Fourier transforms. Related approaches to the solution of linear systems have been used in tomography (see Section IV).

With data acquired in time order, sequential processing is also an option. General approaches to recursive estimation in processing sequences of images are discussed by Green and Titterington [7]. Titterington [8] has provided a recursive EM algorithm for sequential acquisition of data.

The purpose of this paper is to introduce and assess performance of the OS-EM algorithm, and a regularized form (GP in Section V). We aim to show the acceleration of convergence attained with OS. Section II defines OS-EM. Section III discusses choice of subsets and order of processing. Section IV provides a parallel with iterative methods in transmission tomography, particularly MART; EM is a simultaneous form of OS-EM. Section V contains simulation study description and results. Section VI provides discussion. The appendix contains a proof of convergence of OS-EM to a feasible solution with exact projection data.

II. Ordered Subsets EM

A general procedure for sequential processing of data is proposed. While it may be applied with any iterative algorithm, we describe the application with the EM algorithm in SPECT, for concreteness.

Shepp and Vardi's EM algorithm is implemented by iterations (until convergence) requiring projection and back-projection calculations, described below. An initial estimate of image activity is provided to begin this sequence, and the calculations require specification of a projection matrix of weights (often probabilities) providing the link between the unobserved image and the observed projection counts.

With OS-EM, the projection data is grouped in ordered subsets. The OS *level* is defined as the number of these subsets. The standard EM algorithm (i.e. projection followed by back-projection) is then applied to each of the subsets in turn, using the rows of the design matrix corresponding to these ray sums. The resulting reconstruction becomes the starting value for use with the next subset.

We define an *iteration* of OS-EM as a single pass through all the specified subsets. Further iterations may be performed by passing through the same ordered subsets, using as a starting point the reconstruction provided by the previous iteration. With mutually exclusive (and exhaustive) subsets, each OS-EM iteration will have a similar computation time to one standard EM iteration.

Photon recordings on gamma cameras are binned to provide counts y_t on detectors indexed by t. These recordings are the result of emitter activity (modelled as a Poisson point process) within a region. This activity is assumed to

Submitted January, 1993. Revised October, 1993.

Supported by Australian Research Council grants.

H.M. Hudson and R.S Larkin are with Department of Statistics, Macquarie University, NSW 2109, Australia. E-mail: mhudson@zen.efs.mq.edu.au

be uniform within the pixels of a grid imposed on the region; the expected number of photon emissions from pixel j is denoted by x_j. Define the image to be the vector $x = \{x_j : j = 1, \ldots, J\}$. Specify individual pixel and detector elements (or rays) by subscripts j and t respectively. Weights a_{tj} represent the probability that an emission from pixel j is recorded at t. Then detector counts are Poisson distributed with expected values $\mu = Ey = Ax$, where A is the projection matrix. See Shepp and Vardi [2], and Chornboy et al [5] for details of appropriate determination of the weights making allowance for photon attenuation and scatter correction.

Let $\hat{x}^0$ be a prespecified starting image, e.g. uniform. Denote by $\hat{x}^m$ the estimate of x after iteration m. Let $S_1, S_2, \ldots, S_n$ denote the chosen detector subsets in the order selected. The algorithm is then as follows:

1. $m = 0, \hat{x}^m$ initialised, positive
2. Repeat until convergence of $\hat{x}^m$
   (a) $x^1 = \hat{x}^m, m = m + 1$
   (b) For subsets $i = 1, \ldots, n$
      i. *project.*
         Calculate expected values for cumulated counts y as
$$\mu_t^i = \sum_{j=1}^{J} a_{tj} x_j^i \qquad (1)$$
         for detectors $t \in S_i$.
      ii. *backproject.* Calculate
$$x_j^{i+1} = x_j^i \sum_{t \in S_i} \frac{y_t a_{tj}}{\mu_t^i} / \sum_{t \in S_i} a_{tj} \qquad (2)$$
         for pixels $j = 1, \ldots, J$.
   (c) $\hat{x}^m = x^{n+1}$.

Examples:

1. *Non-overlapping subsets*:
   With n subsets each of T detectors per projection, set $S_1 = \{1, 2, \ldots, T\}, S_2 = \{T + 1, \ldots, 2T\}, \ldots, S_n = \{(n-1)T + 1, \ldots, nT\}$, and continue to cycle in turn through each of these subsets. Data used in each sub-iteration comprises counts of all photon scintillations recorded on the first, second, ..., last projection specified. The restoration at stage $(i - 1)$ is modified to incorporate data acquired on projection i.
2. *Cumulative subsets*: Alternatively, set $S_1 = \{1, \ldots, T\}, S_2 = \{1, \ldots, 2T\}, \ldots, S_n = \{1, \ldots, nT\}$. Again, cycle in turn through each of these subsets. Backprojected data from *all* projections already processed is *combined* to form the current restoration of the image. This type of approach was proposed in a more general setting by Titterington [8].
3. *Standard EM*: With n projections of T detectors, employing a single subset $S_1 = \{1, \ldots, nT\}$ provides standard EM.

Notes:

a. We shall refer to this procedure as ordered subsets EM (OS-EM). Should algorithms other than EM be of interest, the backprojection step would be replaced by the corresponding procedure for that algorithm. OS-EM includes standard EM as a particular case, when a single subset includes all projection data.

b. In backprojection, equation (2), the scaling of the image matches the activity expected to be recorded on the projection and the aggregate count recorded on the projection. i.e.
$$\sum_j (\sum_{t \in S_i} a_{tj}) x_j^i = \sum_{t \in S_i} y_t$$
However, initial scaling has no effect on the result of any subsequent EM iteration.

c. If the divisor in equation (2) is zero, set $x_j^{i+1} = x_j^i$.

d. The difference between Vardi-Shepp and ordered subsets EM seems similar, conceptually, to the difference between Jacobi and Gauss-Seidel algorithms for solution of linear equations.

III. Selecting Subsets and Order

In SPECT, subsets may correspond naturally to groups of projections. While the simulations conducted subsequently use subsets corresponding to processing of the projections in opposing pairs, other choices may be considered. For example with three-headed camera systems, in which three projections separated by 120° are recorded simultaneously, it may be useful to use subsets containing three or six projections. We believe it is advantageous to select subsets in a balanced way so that pixel activity contributes equally to any subset.

Since PET collects all tube counts simultaneously, ordered subset EM can best be applied after full collection of counts. The tubes may then be rebinned in parallel families defining projections and our algorithm applied. OS-EM in PET will then provide the same computational advantages in reducing iterations as in SPECT.

The order in which projections are processed is arbitrary, though it may be advantageous to the quality of the reconstruction provided to choose a special order. For example, one might encourage substantial new information to be introduced as quickly as possible by choosing first the projection corresponding to the direction of greatest variability in the image, a second projection perpendicular to the first, and third and fourth projections midway between these, and so on.

IV. Relationship with MART

There are evident similarities between ordered subset EM and the multiplicative algebraic reconstruction technique (MART) of Gordon, Bender and Hermann [9] and related methods. However ordered subsets EM differs in its detail, and is intimately related to the EM algorithm, unlike the other approaches.

Darroch and Ratcliff [10], Censor and Segman [11] and Iusem and Teboulle [12] provide general block-iterative forms of the MART algorithm. OS-EM can also be regarded as a block iterative method. Convergence properties of Censor and Segman's generalization of MART is contained in their Theorem 1. For linear equations for which

there exists a feasible (non-negative) exact solution, then mild conditions on the weights (Section II) and starting value suffice for convergence of block-iterative MART to the maximum entropy solution.

Block-iterative estimates may be contrasted with simultaneous iterative reconstruction techniques, introduced by Gilbert [13]. Simultaneous techniques include all equations relating projection data to parameters simultaneously, whereas MART solves the same equations by introducing subsets of the data in sub-iterations; see Censor [14]. EM can be regarded as a simultaneous form of OS-EM.

EM solves (consistent) non-linear normal equations derived by maximizing the Poisson likelihood function of the data, or similar regularized equations. OS-EM successively increases components of the likelihood corresponding to successive data subsets.

In consistent linear systems (including exact projections) assume there exist feasible solutions to the projection equations $y = Ax$. Maximum entropy and maximum likelihood solutions then satisfy this linear system. OS-EM or block iterative methods provide solutions based on a partition $A^T = [A_1^T, \ldots, A_n^T]$.

OS-EM iterates towards simultaneous non-negative solution of the equations $A_i x = y^i$, the equations for the projections defining block i, for $i = 1, \ldots, n$. We prove, in the Appendix, that under simple conditions on the weights, OS-EM will converge to a feasible solution of the linear system.

EM is known to exhibit linear local convergence with a rate determined the maximum eigenvalue (spectral norm) of the iteration matrix $I - DH$, where D is a diagonal matrix whose components are the solution provided by the iterations, and H is the Hessian of the likelihood function of the full data set. See Green [15], who proves all eigenvalues of this matrix lie in the interval [0,1). In certain cases (strong subset balance) we can compute the local rate of convergence of OS-EM explicitly.

Strong subset balance requires equality of the Hessian (equivalently, Fisher-information) matrices of each data subset's likelihood function; i.e. $H_1 = \cdots = H_n$, and $H = nH_1$. Strong subset balance implies that each subset contains equal information about image activity parameters. Since OS-EM is formed by n EM iterations, it is readily verified that convergence is again linear, with iteration matrix $[I - nDH_n] \ldots [I - nDH_1] = [I - DH]^n$. Strong subset balance provides an iteration matrix whose eigenvalues are precisely the n-th power of the previous eigenvalues. The largest eigenvalue of this matrix is therefore very small in comparison to the largest eigenvalue of EM's iteration matrix. With strong subset balance, convergence of OS-EM will therefore imply linear convergence at a geometrically improved rate (with exponent the number of subsets).

The results above for exact projection data provide some confidence in the convergence of the algorithm in ideal circumstances. With noisy data though, inconsistent equations result. The results of the Appendix are not applicable. While it seems likely from our experiments that OS-EM cycles through a number of distinct limit points, as with MART and many block iterative methods, we conclude this section by providing an argument suggesting that these distinct limit points cannot be very different in many SPECT and PET applications.

OS-EM resolves the linear system determined by A through a partition $A_1, \ldots, A_n$, as above. In the geometry applying to SPECT and PET systems (associated with projection symmetries) this partition has special orthogonality properties.

Subsets should be selected so that $\sum_{t \in S_i} a_{tj}$ is independent of i, so that pixel activity contributes equally to any subset. In SPECT, with subsets comprised of equal numbers of projections, this condition is satisfied if there is no attenuation, or if the subsets chosen provide this balance. We may then assume $\sum_{t \in S_i} a_{tj} = 1$ without essential loss of generality.

Suppose also that the matrix ADA^T is block diagonal for any diagonal matrix D of positive elements. In particular, for $D = \mathrm{diag}(x^i)$, where x^i is the current estimate, assume $A_k D A_i^T = 0$, for $k \neq i$, with $i, k \in \{1, \ldots, n\}$.

Then subiteration i of OS-EM has no effect on fitted values for any projection $k \neq i$, since on substituting for x^{i+1} according to equation (2) we obtain

$$A_k(x^{i+1} - x^i) = A_k D A_i^T z = 0,$$

where z is the vector with components $(y_t - \mu_t^i)/\mu_t^i$, for $t \in S_i$. Hence the Poisson likelihood function for all count data following subiteration i is (apart from a constant)

$$\begin{aligned} L(y, x^{i+1}) &= \sum \{y_t \log \mu_t^{i+1} - \mu_t^{i+1}\} \\ &= \sum_k \sum_{t \in S_k} \{y_t \log \mu_t^{i+1} - \mu_t^{i+1}\} \\ &= L(y, x^i) + \sum_{t \in S_i} \{y_t \log \mu_t^{i+1} - \mu_t^{i+1}\} \\ &\qquad - \sum_{t \in S_i} \{y_t \log \mu_t^i - \mu_t^i\} \end{aligned}$$

The second equality assumes non-overlapping subsets. The third equality follows since the sub-iteration affects only fitted values for subset S_i. But the final two terms of the right hand side provide the increase in the likelihood function of the data subset $\{y_t, t \in S_i\}$ resulting from sub-iteration $i+1$. This sub-iteration applies one standard EM iteration for the model $\mu^i = A_i x$ to this data subset, and hence has the EM property of increasing the likelihood function of this data subset, so that the sum of these terms is positive. In the circumstances above, this implies that the likelihood function of the full set of projection data is also increased, and OS-EM increases this likelihood function within each sub-iteration. By adapting an argument of Vardi, Shepp and Kaufman [16], this implies convergence of the OS-EM algorithm.

The orthogonality condition above cannot apply exactly in tomography, but our experience is that it is adequate as an approximation. In particular it is rare to observe a decrease in the likelihood function in any sub-iteration,

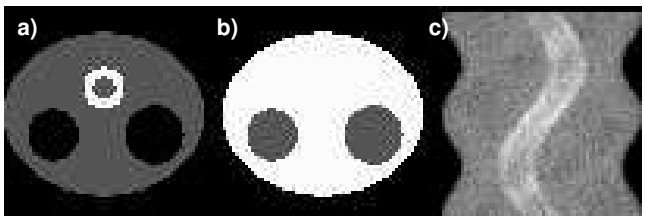

Fig. 1. Computer Simulated Chest Cross-section: (a) chest phantom emitter activity; (b) attenuation map; (c) sinogram.

and we have never observed a decrease in likelihood over a complete iteration of OS-EM. An argument, given in Barnett et al. [17], for a similar orthogonality condition based on the geometry of the detectors is applicable here, too.

V. Simulation Study

To study the properties of ordered subsets we conducted simulation studies employing computer generated projection data based on a model ("chest phantom") for activity. Generation incorporated attenuation and Poisson noise. 64 projections were generated over 360°. Counts were recorded in 64 bins per projection. Counts recorded on all projections totalled approximately 410 thousand. Figure 1 shows the chest phantom activity and attenuation map, and the simulated projection data (sinogram). Chest phantom activity is concentrated in a ring of high activity ("myocardium"). There are two regions of very low activity ("lungs") with otherwise uniform low activity within an elliptical body region, of cross section 40 x 32 cm. The activity in myocardium, background and lungs were specified to be in the ratio 8:1:0. Attenuation coefficients were 0.03/cm in low activity ("lung") regions, 0.12/cm elsewhere within the body ellipse, and 0.000/cm outside the body.

Ordered subsets were applied with two algorithms, providing OS-EM and OS-GP algorithms. OS-EM is the adaption of Shepp-Vardi EM described in Section II. OS-GP is the OS adaption of Green's one-step-late (OSL) reconstruction.

GP provides MAP estimation based on Gibbs priors, for which a penalized likelihood criterion is maximized. This criterion is

$$L^*(x) = \sum [y_t \log \mu_t(x) - \mu_t(x)] - \beta \sum_{s,r} w_{sr} \phi(\frac{x_s - x_r}{\sigma}),$$

where β and σ are parameters of the procedure, and ϕ is a log-cosh function which penalizes discrepancies between pixel neighbours s, r in a manner determined by the fixed weights w. Maximizing the criterion function L^* balances two objectives: increasing the likelihood function (the first sum of the expression); and, reducing the roughness penalty (the second sum).

GP was specified with and defined as in Green [4], with parameters β =0.006, σ = 2.00. These parameter values had been established as suitable for standard Gibbs prior reconstructions of chest phantom data.

OS-EM was defined by equations (1) and (2), using non-overlapping subsets. Results for cumulative subsets are not reported here because of the similarity of its results with those of standard EM.

The variants of OS-EM and OS-GP used in the simulation are distinguished by a trailing number indicating the OS level (number of subsets). Levels considered were 1 (standard EM or OSL), 2, 4, 8, 16 and 32. Note that all variants of OS-EM take equal time to compute a single iteration.

Ordering of subsets was designed to introduce independent information about the image in successive subsets. The sequence used with OS-EM level 32 introduced projection pairs in the order $0°, 90°, 45°, 135°, 22.5°, 112.5°, 67.5°$ etc.

The reconstruction was scaled after each iteration so that total expected counts agreed with total counts (a property of the ML solution). This step is unnecessary; it has no effect on qualitative appearance of the reconstruction, or indeed on the subsequent iterations, as rescaling by any constant multiplier has no effect on EM. It was applied to provide consistency of scaling with ML, so that criteria used to compare solutions were not affected by simple scaling effects.

The scaling may be conducted as follows. The image $\hat{x}$ obtained at step 2(c) of the OS-EM definition in Section 2 is replaced by $c\hat{x}$, where $c = (\sum_t y_t)/(\sum_j \hat{x}_j a_{.j})$, with $a_{.j} = \sum_t a_{tj}$. Projection weights a_{tj} were precalculated for computational efficiency in our code. As a consequence the weights $a_{.j}$ were available, and scaling was a trivial operation.

Chi-square and MSE measures of error were then calculated. Chi-square (otherwise known as deviance) is defined as $G = 2\sum[y_t \log(y_t/\mu_t) - (y_t - \mu_t)]$, where $\{\mu_t\}$ are the fitted projections. Chi-square measures discrepancy between fitted projections provided by a reconstruction and the counts on all projections; MSE, defined as $\sum(\hat{x}_j - x_j)^2/J$, measures an average discrepancy between a reconstruction and the chest phantom image. Note that, for given data y, G differs from the Poisson likelihood only by a fixed constant. Therefore chi-square decrease ensures a corresponding likelihood function increase in the context of this study.

Figure 2 compares the reconstructions after a single iteration of each of the OS-EM variants. It can be seen clearly that higher OS levels provide better definition for small numbers of iterations, levels 16 and 32 providing low MSE after a single iteration.

Figure 3 compares the reconstructions after matched iterations of each of the OS-EM variants. The reconstructions have very similar chi-square and mean square error and are visually similar. Thirty two iterations of standard EM provide similar results to one OS-EM (level 32) iteration.

Figure 4 shows two cross sections, one centrally through the ring of high activity (myocardium), the other through the regions of low activity (lungs). A single EM iteration provides poor estimates of high or low activity. OS-EM-32, after one iteration, closely matches the true activity section (solid line) and the reconstruction from 32 iterations of

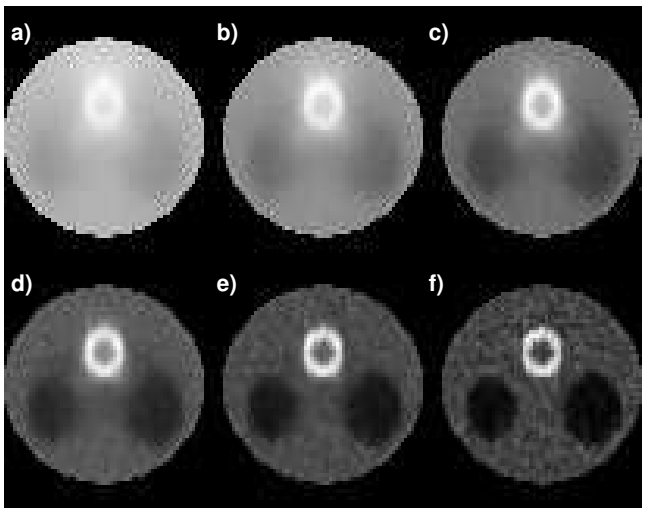

Fig. 2. OS-EM reconstructions after one iteration, by level: 1, 2, 4 (a-c); 8, 16, 32 (d-f).

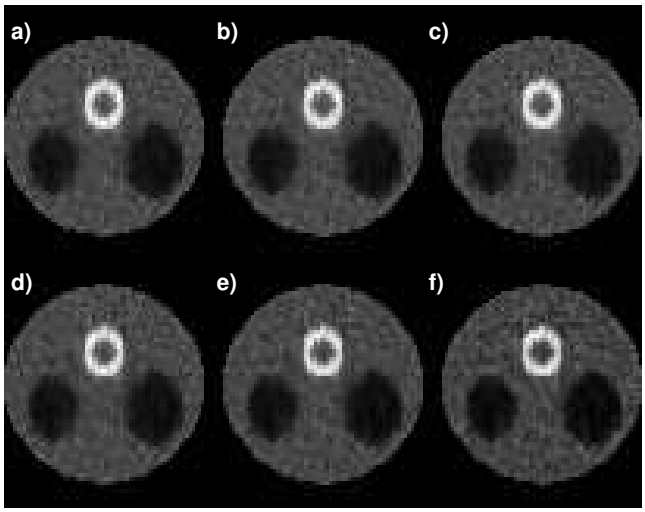

Fig. 3. OS-EM comparison after matched iterations: (a) level 1, 32 iterations; (b) level 2, 16 iterations; (c) level 4, 8 iterations; . . .; (f) level 32, 1 iteration.

standard EM.

Figure 5 plots chi-square versus iteration number in the first 50 iterations of all OS-EM variants. While standard EM is slow to converge, OS-EM converges quickly at levels 16 and 32, with only two iterations required to produce a lower chi-square than that obtained by EM after 50 iterations.

Convergence required fewer iterations for higher levels of data subdivision. The lack of fit to observed data, as measured by chi-square of the ultimate reconstruction, is slightly worse with 32 levels of subdivision than with lower levels. This suggests there is an effective limit to the degree of subdivision that is desirable.

Figure 6 plots mean squared error versus chi-square for 128 iterations of OS-EM-1 and OS-EM-2 (also for OS-GP-1 and OS-GP-2). Note that every OS-EM-1 iteration reduces chi-square, so that the iterations commence on the right of the graph. It can be seen that OS-EM-2 closely matches

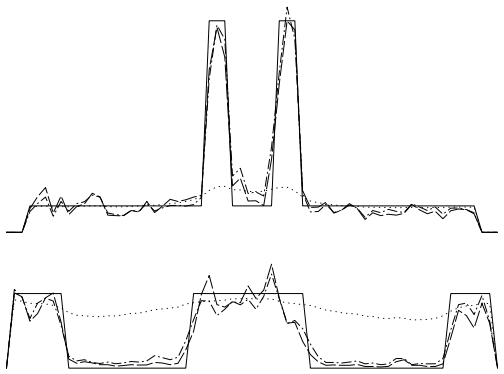

Fig. 4. Traces through myocardium and lung regions. Shown in each case is emitter activity: true phantom (solid line); EM, 1 iteration (dotted) and 32 iterations (dash-dot); OS-EM 32, 1 iteration (broken dashes).

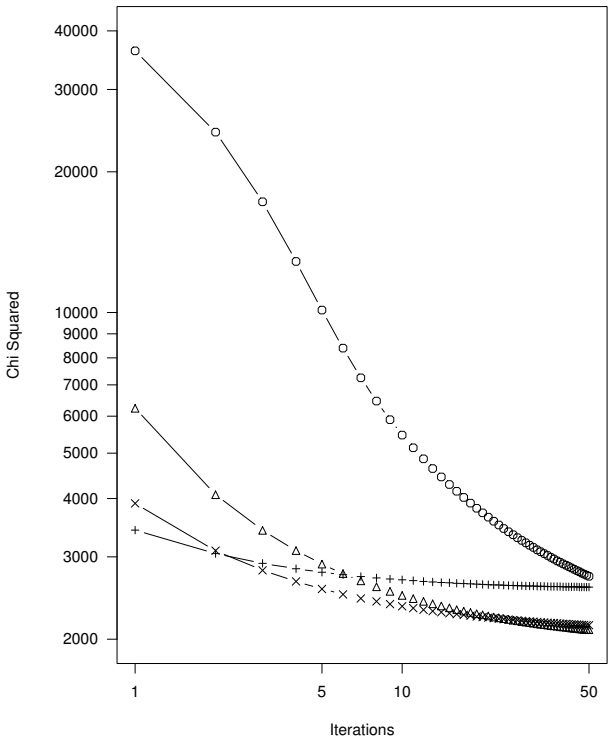

Fig. 5. Chi-square (G) vs Iteration (log scale) by OS-EM level: OS-EM level 1 (○); level 8 (△); level 16 (×); level 32 (+).

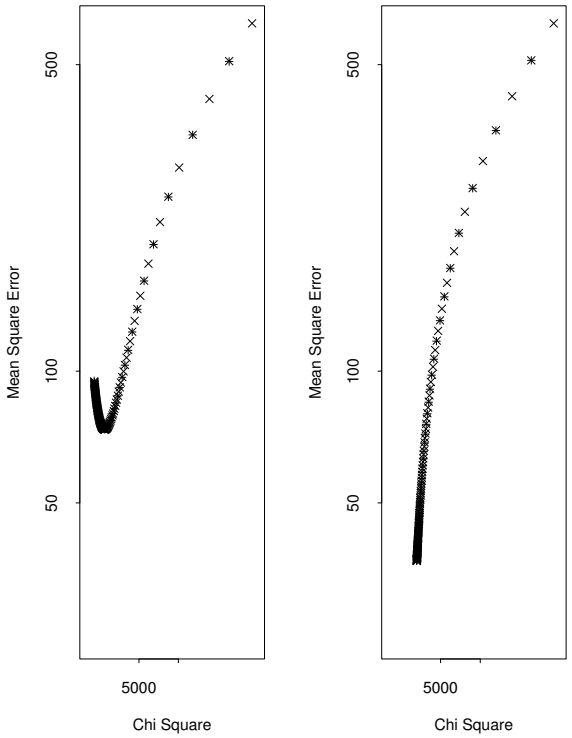

Fig. 6. Mean square error vs Chi-square for OS-EM level 1 (×) and level 2 (+)

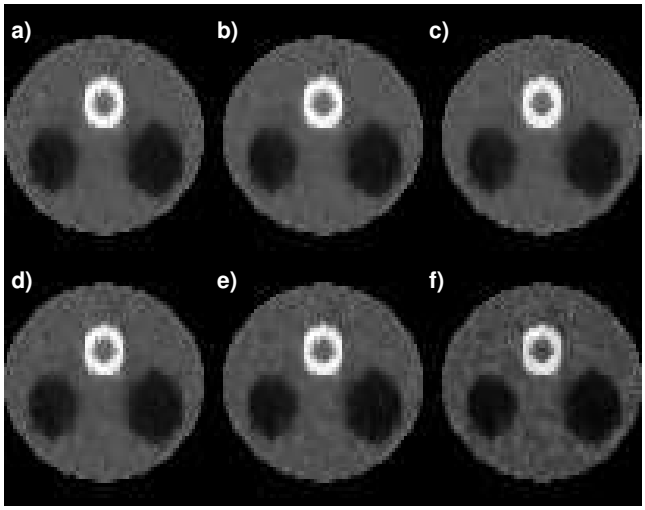

Fig. 7. OS-GP iterations after matched iterations. Reconstructions are: (a) level 1, 32 iterations; (b) level 2, 16 iterations; (c) level 4, 8 iterations; ...; (f) level 32, 1 iteration.

every second iteration of OS-EM-1. This pattern applies at all levels of OS-EM, and also with OS-GP iterations.

Now examine reconstruction quality after large numbers of iterations. Figure 6 displays a turnaround in MSE, caused by an increasing noise artifact in EM reconstructions. This is a well known phenomenon when the EM algorithm is applied to noisy projection data. GP effectively eliminates this noise artifact; this entails a slight sacrifice in fit to observed data.

Figure 7 compares the reconstruction after matched iterations of OS-GP. Again, very similar high definition results are obtained from high levels of OS-GP. It should be noted though that, due to one step late calculation overhead, OS-GP-32 computation time per iteration is twice that of OS-GP-1 (GP). Nevertheless, a substantial acceleration remains.

Figure 8 illustrates this acceleration. Again OS, at levels 32 and 16, provides better fit to projection data after one or two iterations than the OSL procedure, after 32 iterations. The MSE of OS-GP-32 after 1 iteration (MSE=62.5) is less than the MSE for GP after anything up to 32 iterations (MSE=68.7 after 32 iterations).

Figure 9 displays the limit points of OSEM and OSGP algorithms, by plots of MSE and chi-square at selected iterations. Again, successive iterations reduce chi square, so iterations commence on the right of each graph. The noise

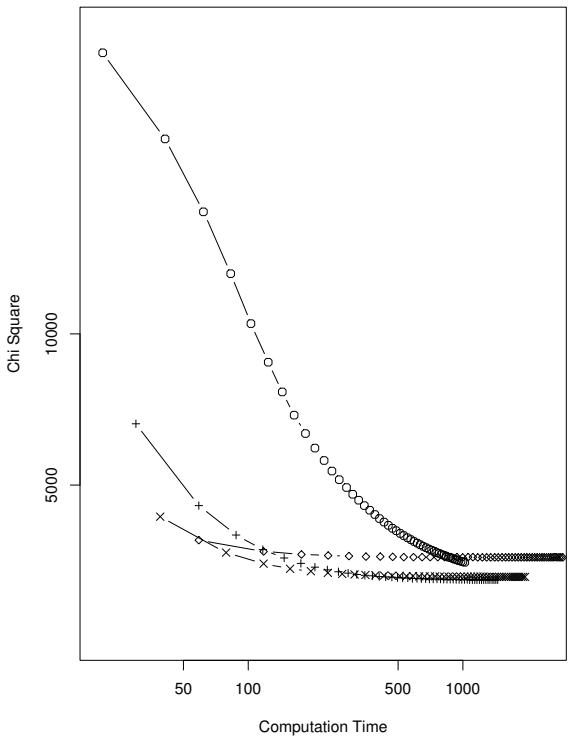

Fig. 8. Chi-square versus computation time for 50 iterations of OS-GP. Levels are: 1 (◯); 8 (+); 16 (×); 32 (◇).

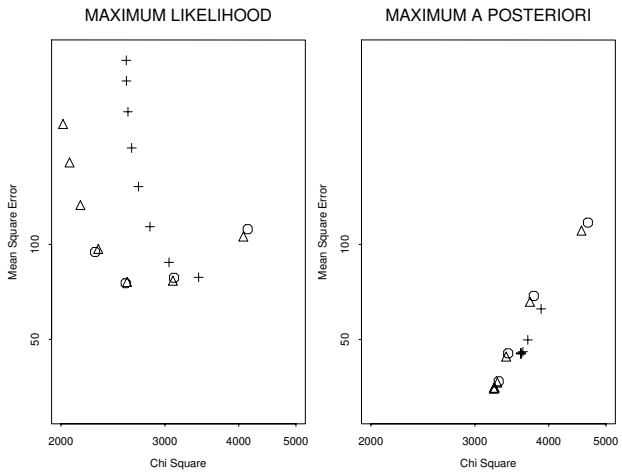

Fig. 9. Convergence of OS-EM and OS-GP. Mean square error versus chi-square at selected iterations: level 1, iterations 16,32,64,128 (○); level 8, iterations 2, 4, 8, ..., 128 (△); level 32, iterations 1, 2, 4, ..., 128 (+).

artifact with OS-EM-32 is greater than with OS-EM-1 or OS-EM-8. Similarly OS-GP-32 converges to a limit different from the OS-GP-1 limit. However the OS-GP-8 limit and OS-GP-16 limit (not shown, to enhance clarity) are for all practical purposes identical with the OS-GP-1 limit, whose MSE is far superior to that of OS-EM-1.

All simulations were run on an 80486 processor running at 33 Mhz and programmed in Turbo Pascal 5.5. A single iteration of any of the variants took no more than 40 seconds.

VI. Discussion

The simulation results above clearly indicate ordered subsets (OS) methods provided an order improvement in speed of execution in this example. To achieve a reconstruction with a given maximum level of error, the number of iterations required were inversely proportional to the level of OS. As projection data was subdivided into more subsets, accelerated reduction in chi-square occured up to a critical number of iterations. Beyond this critical number, the noise artifact of ML solutions is magnified by OS-EM (but not by OS-GP). This well known noise artifact motivates stopping short iterations of Vardi-Shepp EM.

Rather than stopping short EM iterations to obtain optimal mean square error, the same MSE is obtainable with small numbers of OS iterations.

The results reported are typical of those from many other simulation studies we have conducted. Further clinical studies are ongoing in collaboration with Department of Nuclear Medicine, Royal Prince Alfred Hospital (see Hudson, Hutton and Larkin, [18]).

It seems clear that, except with noise free data, OS-EM converges to a non-ML solution. The questions of whether OS-EM necessarily converges, and if so, to what limit, remain open. Empirically, it appears that OS-EM follows a cycle between different limit images depending on the last projection processed. Such behaviour would match the theoretical properties of the ART algorithms applicable with linear systems. Our simulations suggest that in real scale applications the different limit images are similar, and indeed, with moderate numbers of subsets, indestinguishable.

If a ML solution is required, a composite method may be defined in which a single iteration comprises successive application of OS-EM at a sequence of levels descending to level 1. For example, one iteration of the composite method could employ levels 32, 16, 8, 4, 2, 1. Our computational experience is that this algorithm increases the likelihood function at each iteration, and hence convergence to a ML solution is attained. At the same time, overall computations are greatly reduced.

Without such adaptions, the most appropriate level of data subdivision depends on a number of factors. These are attenuation density, subset balance, and level of noise in projection data.

With no attenuation, use of individual projections as subsets is inefficient, as opposite pairs of projections provide estimates of the same ray sum, and should be included together in the same subset.

Subset balance is a second factor. In our simulations, subset imbalance (variability in the probability of a pixel emission being detected in different subsets) will be greatest for subsets composed of an individual projection, but balance will improve as the number of subsets is reduced.

With 64 projections of data available, both factors motivated our use of at most 32 data subdivisions. With fewer subdivisions slightly better likelihood values are eventually obtained, as observed in the comparison of OS-EM-32 and OS-EM-16 in Figure 5. The discrepancy increases with level of statistical noise in projections, so that with low counts we prefer to use less data subdivision, e.g. 4 projections per subset instead of 2.

The OS method shows great promise for practical use. OS-EM offers the benefits of EM (quality of restoration) without the computational burden that may have hindered its acceptance. An order reduction in computations will assist in making real time processing of SPECT and PET data viable with current technologies.

As a single iteration of OS-EM-32 provides a reasonable reconstruction, it is even possible to process projection data immediately as it is acquired. As data collection is often time consuming, this effectively removes all post-processing overhead in providing reconstructions. Such an approach may find use in medical studies of dynamic systems.

Ordered subset EM is also suited to an arbitrary sequence of projections, as might be obtained for example when patient movement is detected during the acquisition of data, as the coefficients represented in the design probability matrix, A, can be computed in sequence, as required. See Fulton, et al, [19].

While our discussion has been limited to ordered subset variants of the EM algorithm, OS methods can be readily applied with algorithms other than EM. The basic approach advocated is flexible and readily modified for use with other restoration algorithms and for real time restora-

tions with other data collection schemes (camera technologies).

Acknowledgement

Brian Hutton, and colleagues in the Department of Nuclear Medicine, Royal Prince Alfred Hospital, have assisted in this formulation and added much to our understanding of SPECT. Our thanks to Jeanne Young who introduced us to the links with ART and SIRT. Also to Adrian Baddeley, Jun Ma and Victor Solo for helpful discussions. This paper was prepared in part while Hudson was a visitor to Division of Mathematics and Statistics, CSIRO.

References

[1] P.P.B. Eggermont, G.T. Herman, and A. Lent, "Iterative algorithms for large partitioned linear systems, with applications to image reconstruction", *Linear Algebra and Its Applications*, vol. 40, pp. 37–67, 1981.

[2] L.A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography", *Transactions on Medical Imaging*, vol. MI-2, pp. 113–122, 1982.

[3] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-d Bayesian reconstruction from Poisson data using Gibbs priors", *IEEE Transactions on Medical Imaging*, vol. MI-8, pp. 194–202, 1989.

[4] P.J. Green, "Bayesian reconstruction from emission tomography data using a modified EM algorithm", *IEEE Transactions on Medical Imaging*, vol. MI-9, pp. 84–93, 1990.

[5] E.S. Chornboy, C.J. Chen, M.I. Miller, T.R. Miller, and D.L. Snyder, "An evaluation of maximum likelihood reconstruction for SPECT", *IEEE Transactions on Medical Imaging*, vol. MI-9, pp. 99–110, 1990.

[6] L Kaufman, "Implementing and accelerating the EM algorithm for Positron Emission Tomography", *IEEE Transactions on Nuclear Medicine*, vol. MI-6, pp. 37–51, 1987.

[7] P.J. Green and D.M. Titterington, "Recursive methods in image processing", *Proc. 46th Session ISI*, 17 pages, 1990.

[8] D.M. Titterington, "Recursive parameter estimation using incomplete data", *J. Royal Statist. Soc., B*, vol. 46, pp. 257–267, 1984.

[9] R. Gordon, R.and Bender and G.T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography", *J. Theor. Biol.*, vol. 29, pp. 471–481, 1970.

[10] J. N. Darroch and D. Ratcliff, "Generalized iterative scaling for log-linear models", *Annals Math. Statist.*, vol. 43, pp. 1470–1480, 1972.

[11] Y. Censor and J. Segman, "On block-iterative entropy maximization", *J. Information and Optimization Sciences*, vol. 8, pp. 275–291, 1987.

[12] A.N. Iusem and M Teboulle, "A primal-dual iterative algorithm for a maximum likelihood estimation problem", *Computational Statistics and Data Analysis*, vol. 14, pp. 443–456, 1992.

[13] P. Gilbert, "Iterative methods for the three-dimensional reconstruction of an object from its projections", *J. Theor. Biol.*, vol. 36, pp. 105–117, 1972.

[14] Y. Censor, "Finite series-expansion reconstruction methods", *Proc. IEEE*, vol. 71: 3, pp. 409–419, 1983.

[15] P.J. Green, "On use of the EM algorithm for penalized likelihood estimation", *J. Royal Statist. Soc., B*, vol. 52: 3, pp. 443–452, 1990.

[16] Y. Vardi, L.A. Shepp and L. Kaufman, "A statistical model for Positron Emission Tomography (with discussion)", *J. American Statist. Soc.*, vol. 80, pp. 8–20, 1985.

[17] G. Barnett, S. Crowe, M. Hudson, P. Leung, K. Notodiputro, R. Proudfoot, and J. Sims, "The use of small scale prototypes in image reconstructions from projections", *J. Applied Statistics*, vol. 16, pp. 223–242, 1989.

[18] H.M. Hudson, B.F. Hutton, and R. Larkin, "Accelerated EM reconstruction using ordered subsets", *J. Nucl. Med.*, vol. 33, (abs), pp. 960, 1992.

[19] R.R. Fulton, B.F. Hutton, M. Braun, B. Ardekani, and R.S. Larkin, "Use of 3d reconstruction to correct for patient motion in SPECT", *Phys. Med. Biol.*, vol. 39, 3, pp. 563-574, 1994.

Appendix

Proof of Convergence of OS-EM with subset balance and exact projections

Assumptions

(A1) Assume the existence of some feasible solution $x^* > 0$ to the linear system $y = Ax$. We assume all $a_{tj} \geq 0$, and that for each at least one of these coefficients is strictly positive. By setting appropriate components of x to zero, we may reduce the parameter set when $y_t = 0$ for any index t, so without loss of generality assume $y > 0$.

Under assumption (A1) we may define iterations x^k by: $x^0 > 0$;

$$\begin{aligned} x_j^{k+1} &= x_j^k \sum_{t \in S_k} \frac{y_t a_{tj}}{\mu_t^k} \\ \mu_t^k &= \sum_j a_{tj} x_j^k \end{aligned} \tag{3}$$

where, at each iteration k, some subset, S_k, of the (finite) index set of the data has been specified.

(A2) Subset balance. Require that, for each subset S_i used in (3), $\sum_{t \in S_i} a_{tj} = 1$, for each j. Let $L^k := L(x^k; x^*)$

where

$$L(x; x^*) = -\sum x_j^* \log \frac{x_j^*}{x_j} \tag{4}$$

Note that L^k is well defined since $x^k > 0$, $x^* \geq 0$, on defining $0 \log 0 = 0$.

Lemma 1 (Count preservation)

1. For $k \geq 0$, $\sum x_j^{k+1} = \sum_{t \in S_k} y_t$
2. (A2) implies $\sum_{t \in S_k} \mu_t^k = \sum x_j^k$ and $\sum_{t \in S_k} y_t = \sum x_j^*$, independent of $k \geq 1$.
3. Given $y > 0$, $\mu_t^k > 0$ for all t, and (A2), the deviance

$$\sum_{t \in S_k} y_t \log(\frac{y_t}{\mu_t^k}) \geq \sum_{t \in S_k} y_t (1 - \frac{\mu_t^k}{y_t}) = 0$$

Proof:

$$\begin{aligned} \sum x_j^{k+1} &= \sum x_j^k \sum_{t \in S_k} \frac{y_t a_{tj}}{\mu_t^k} \\ &= \sum_{t \in S_k} \frac{y_t}{\mu_t^k} \sum_j a_{tj} x_j^k \\ &= \sum_{t \in S_k} y_t \\ \sum_{t \in S_k} \mu_t^k &= \sum_{t \in S_k} \sum_j a_{tj} x_j^k \end{aligned}$$

$$= \sum_j x_j^k \sum_{t \in S_k} a_{tj}$$
$$= \sum x_j^k$$

The third equality follows substituting x^* for x^k in the above.

The final inequality follows from the inequality $\log x \geq 1 - x^{-1}$, for $x > 0$. ■

Proposition 1: Under assumptions (A1) and (A2), with $\{x^k\}$ defined in (3), $\{L^k\}$ is nondecreasing and bounded above.

Proof: Let

$$\begin{aligned}
\Delta^k &= L^{k+1} - L^k \\
&= \sum x_j^* \log(x_j^{k+1}/x_j^k) \\
&= \sum_j x_j^* \log(\sum_{t \in S_k} \frac{y_t a_{tj}}{\mu_t^k}) \\
&\geq \sum_j x_j^* \sum_{t \in S_k} a_{tj} \log(y_t/\mu_t^k) \\
&= \sum_{t \in S_k} \log(y_t/\mu_t^k) \sum_j a_{tj} x_j^* \\
&= \sum_{t \in S_k} y_t \log(y_t/\mu_t^k) \\
&\geq 0 \qquad (5)
\end{aligned}$$

Here the third equality uses the definition of OS-EM, the first inequality follows by application of Jensen's inequality, noting assumption (A2), and the final inequality applies the Lemma.

L^k is bounded above by 0, applying the inequality $\log x \geq 1 - x^{-1}$ to (4), again noting the count preservation. ■

From Proposition 1, $\{L^k\}$ converges, so $\Delta^k \to 0$. But this implies the *deviance* $\sum_{t \in S_k} y_t \log(y_t/\mu_t^k) \to 0$, by (5). This is sufficient to prove convergence of OS-EM, as provided by the result:

Proposition 2: With assumptions (A1), (A2) and algorithm (3):

1. For any subset T_i used infinitely often ("i.o.") in the sequence $\{S_k\}$, $\mu_t^k \to y_t$ for $t \in T_i$ (i.e. convergence of fitted values occurs for $t \in T_i$) along any subsequence of iterations using T_i exclusively;

2. If subsets are selected from a stock $\{T_1, \ldots, T_n\}$ in such a way ("nearly cyclic control") that $\exists N$ for which the set $\{S_k, \ldots, S_{k+N}\}$ contains (for any k) all members of the stock then $x^k \to \tilde{x}$, where $\tilde{x}$ is a feasible solution of the linear system for $\{y_t : t \in T = \cup_{i=1}^n T_i\}$.

Proof: Consider the subsequence K_1 of integers, indicating iterations using a particular subset, say T_1, occurring i.o. Then by the remark following Proposition 1, the deviance *for this subset*, $\sum_{t \in T_1} y_t \log(y_t/\mu_t^k) \to 0$ for $k \in K_1$. But the condition for equality in (5), and continuity of the deviance function, together imply $\mu_t^k \to y_t$ for $k \in K_1$, for all $t \in T_1$.

We now show $\mu_s^k \to y_s$ for $k \in K_1$, for all $s \in T$, i.e. convergence occurs for *all* the data subsets.

Assume subsets S_k are selected from the stock $\{T_1, \ldots, T_n\}$, and each element of the stock is selected i.o. Then, by the argument just given, $\mu_t^k \to y_t$, for all $t \in T_i$, along any subsequence using T_i exclusively. Hence, for $k_1 \in K_1$ and $s \in T_i$, for $i \in \{2, \ldots, n\}$

$$\begin{aligned}
&|\sum_j a_{sj} x_j^{k_1} - y_s| \\
&= |\sum_j a_{sj} x_j^{k_i} \prod_{k=k_i}^{k_1-1} (\sum_{t \in S_k} \frac{y_t a_{tj}}{\mu_t^k}) - y_s| \\
&= |\sum_j a_{sj} x_j^{k_i} \exp\{\sum_{k=k_i}^{k_1-1} \log(\sum_{t \in S_k} \frac{y_t a_{tj}}{\mu_t^k})\} - y_s| \\
&\leq |\sum_j a_{sj} x_j^{k_i} - y_s| \\
&\qquad + |\sum_j a_{sj} x_j^{k_i} \{\exp[\sum_{k=k_i}^{k_1-1} \log(\sum_{t \in S_k} \frac{y_t a_{tj}}{\mu_t^k})] - 1\}|
\end{aligned}$$

where k_i is the largest integer less than or equal to k_1 such that $S_{k_i} = T_i$. Since all subsets occur i.o. the first term approaches 0 as k_1 – and hence k_i – approaches ∞. Assuming nearly cyclic control, there are at most N terms in the sum of logarithms, and each term approaches 0 since $\mu_t^k \to y_t$, for all $t \in S_k$. Hence the RHS of the inequality approaches 0 along subsequence K_1, completing the proof that $\mu_s^k \to y_s$ for $k \in K_1$, for all $s \in T_i$.

Consider a point of accumulation $\tilde{x}$ of the subsequence $\{x^k : k \in K_1\}$. Since every member of the sequence belongs to the closed bounded region $\{x : x \geq 0, \sum x_j = \sum x_j^*\}$, $\tilde{x}$ exists and belongs to this region, and is feasible. There exists a subsequence of K_1 along which $x^k \to \tilde{x}$. Then along this subsequence $\mu_s^k \to \lim \sum_j a_{sj} x_j^k = \sum_j a_{sj} \tilde{x}_j$, and by the result above this limit along subsequence K_1 is y_s, for $s \in T$. Hence the limit $\tilde{x}$ is a solution of the linear system for $t \in T$. But then $\lim_{k \to \infty} L(x^k, \tilde{x})$ – which exists by Proposition 1 – is the limit along the subsequence, evaluated as $L(\tilde{x}, \tilde{x}) = 0$. By the continuity of L, $\tilde{x}$ is the only point of accumulation of x^k, implying convergence to $\tilde{x}$. ■

Corollary 2.1: The OS-EM algorithm with subsets selected from a fixed stock of exhaustive (but possibly overlapping) subsets of the index set of the data by cyclic control (as in Section II) or by almost-cyclic control converges, under (A1) and (A2), to a feasible solution of the full linear system.